

Conference Program

JANUARY 12-14, 2024 | MACAU, CHINA | GMT+8

2024 6th International Conference on Image
Processing and Machine Vision (IPMV 2024)

2024 5th International Conference on Electronics
and Signal Processing (ICESP 2024)

Published by:



Co-sponsored by



Technical Supported by



Supported by



Hotel Royal Macau

Address: Estrada da Vitoria No.2-4, Macau

Table of Content

General Information	02
Welcome Message.....	04
Conference Committee	05
Agenda Overview	06
Keynote Speaker	08
Onsite Session 1A: Data Communication and Image Processing Technology	11
Onsite Session 1B: Data Communication and Image Processing Technology	14
Online Session 1: Image Detection and Processing.....	17

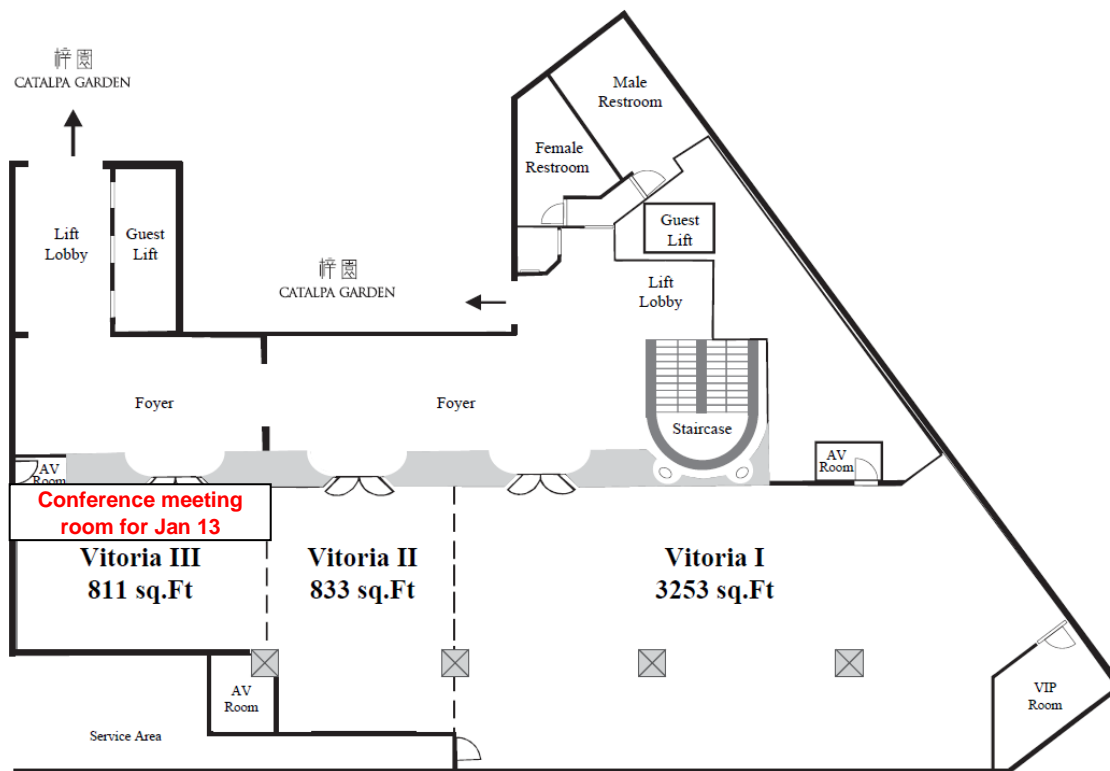
Note

General Information

1)

Conference Venue: Hotel Royal Macau

Address: Estrada da Vitoria No.2-4, Macau



Friday, Jan 12, 2024: Reception table set in the Lobby

Saturday, Jan 13, 2024: Vitoria III, 2nd floor

Access to Hotel Royal Macau

Transportation:

Travelling around Macau is easy with a choice of convenient and safe transport options, and many great heritage sites are within walking distance from Hotel Royal Macau. The hotel room guests can also exclusively enjoy complimentary shuttle service commuting between the hotel and Macau Outer Harbour Ferry Terminal and Border Gate (please check the timetable below for details).

Besides, buses and taxis are the most common forms of public transport. For buses, you can either pay by cash or a stored value smart card, Macau Pass, which can be bought, topped up and returned at most convenient stores in town. This will not only avoid the hassle of getting enough coins for each ride, but also save the cost of multiple travel on the same day. Our friendly Concierge team is always at your service to update about the latest local transportation tips.

2) Onsite Registration

Registration desk (Reception table set in the lobby of Hotel Royal Macau) → Inform the staff of your paper ID → Sign-in → Claim your conference kit.

3) Devices Provided by the Organizer

Laptops (with MS-Office & Adobe Reader) / Projectors & Screen / Laser Sticks

4) Materials Provided by the Presenter

Oral Session: Slides (pptx or pdf version). Format 16:9 is preferred.

Official language: English.

5) Duration of Each Presentation

※ Keynote Speech: 45min, including Q&A / Oral Presentation: 15min, including Q&A


6) Notice

※ Please wear your delegate badge (name tag) for all the conference activities. Lending your participant card to others is not allowed.

※ Please take good care of your valuables at any time during the conference. The conference organizer does not assume any responsibility for the loss of personal belongings of the participants during conference day.

※ **UTC+8. Please be aware of time difference between this and your region / country.**

7) Online Presentation Tips

 Zoom Download	Meeting ID	Link
	ZOOM ID: 880 7857 6196	https://us02web.zoom.us/j/88078576196

Note:

We recommend that you install the Zoom platform on your computer before the conference starts. New users can participate in the Zoom meeting without registration.

Participants who are going to do an online presentation are required to join the rehearsal in Zoom on Friday, January 12. Duration: 3min apiece. Feel free to leave after you finish the test.

◆ Name Setting

Keynote Speaker: KN-Name

Committee: Position-Name

Author: Paper ID-Name

Listener: Listener-Name

◆ Useful Links

◇ [Conference Banner](#)

◇ [Zoom Background](#)

Welcome Message

Dear researchers, delegates of conference,

Welcome to Macau, China to attend 2024 6th International Conference on Image Processing and Machine Vision (IPMV 2024) and 2024 5th International Conference on Electronics and Signal Processing (ICESP 2024) which is co-sponsored by University of Macau.

In consideration of health and safety for everyone and travel restrictions, IPMV 2024 and ICESP 2024 are still made offline and online mixed. We feel pity that we cannot gather all together in Macau. However, we will work hard to provide you with a high-quality conference as always, and with the hope that we can get your support.

The objective of the conferences is to provide a premium platform to bring together researchers, scientists, engineers, academics and graduate students to share up-to-date research results. We are confident that during this time you will get the theoretical grounding, practical knowledge, and personal contacts that will help you build a long term, profitable and sustainable communication among researchers and practitioners in the related scientific areas.

This year we have 3 Keynote Speeches. They are Prof. James Tin-Yau Kwok, from The Hong Kong University of Science and Technology, Hong Kong, China; Prof. Kenneth K.Y. Wong, from University of Hong Kong, Hong Kong, China and Prof. Seokwon Yeom, from Daegu University Gyeongsan, South Korea; We have 2 sessions (In person + Virtual) with topics: Image Detection and Processing, Data Communication and Image Processing Technology

Meanwhile, we received more than 60 submissions from research institutions, universities and industries. The papers in the proceedings are accepted after being peer-reviewed by conference committee, international reviewers based on the topic and quality. With the keynote speeches, oral sessions, we'll have an exciting program this year, which will allow participants to present and discuss the latest research and industrial developments in these fields.

On behalf of the organizing committee, we would like to deeply express our heartfelt appreciation to all our delegates, keynote speakers, session chairs, as well as all the committee members involved in the technical evaluation of conference papers and in the organization of the conference for their time, effort, and great contributions.

We also wish that this conference will be an unforgettable and wonderful experience for you.

With Warmest Regards,

Conference Organizing Committees

IPMV 2024

Ms. Ching Cao

Email: ipmv_conf@yeah.net

ICESP 2024

Ms. Fiona

Email: icesp_conf@163.com

And you can also add the following WeChat Assistant



Conference Committee

Conference Advisory Committee

Fei Richard Yu, Carleton University, Canada (IEEE Fellow)

Conference General Chair

Chi Man Pun, University of Macau, Macau, China

Conference General Co-Chair

James Tin-Yau Kwok, The Hong Kong University of Science and Technology, Hong Kong, China (IEEE Fellow)

Conference Organizing Co-Chair

Kenneth K.Y. Wong, University of Hong Kong, Hong Kong, China

Conference Program Chairs

Hong Lin, University of Houston-Downtown, USA

Hui Zhang, BNU-HKBU United International College, China

Conference Program Co-Chairs

Seokwon Yeom, Daegu University Gyeongsan, South Korea

Xiao Zhou, Wuhan University of Technology, China

Conference Publicity Chair

Kai Han, The University of Hong Kong, Hong Kong, China

Technical Program Committees

Chao Yang, Southeast University, China

Yew Kee WONG Eric, Hong Kong Chu Hai College, Hong Kong, China

Karthick Prasad Gunasekaran, Amazon Web Services Inc, USA

Bin Xue, National University of Defense Technology, China

Wanwan Li, University of South Florida, USA

Xiaochun Cheng, Middlesex University, UK

Shuai Zhao, Jinan University, China

Nusrat Sharmin, Military Institute of Science Technology, Bangladesh

Emerson Nobuyuki Itikawa, Federal University of Goiás, Brazil

Xiantao Jiang, Shanghai Maritime University, China

Suphongsak Khetkeeree, Mahanakorn University of Technology, Thailand

Thaweesak Yingthawornsuk, King Mongkut's University of Technology Thonburi, Thailand

Vinh Truong Hoang, Ho Chi Minh City Open University, Vietnam

Luis Gómez, Universidad de Las Palmas de Gran Canaria, Spain

Aruna Bhat, Delhi Technological University, India

Ruey-shun Chen, National Chiao Tung University, Taiwan, China

Ashwani Kumar Aggarwal, Sant Longowal Institute of Engineering and Technology, India

Lu Tang, Southeast University, China

Chawalit Benjangkaprasert, King Mongkut's Institute of Technology Ladkrabang, Thailand

Lorna Uden, Staffordshire University, UK

Pavlo Maruschak, Ternopil Ivan Puluj National Technical University, Ukraine

Juan Suardiáz Muro, Universidad Politécnica de Cartagena, Spain

Azizi Abdullah, University Kebangsaan Malaysia, Malaysia

Sompong Liangrocpart, Mahanakorn University of Technology, Thailand

Sheli Sinha Chaudhuri, Jadavpur University, India

Ying-Hao Yu, National Chung Cheng University, Taiwan, China

Ghufuran Ahmed, National University of Computer & Emerging Sciences (NUCES), Pakistan

Agenda Overview (UTC+8)


Friday, January 12, 2024		
Onsite Registration	14:00-17:00	Reception table set in the lobby of Hotel Royal Macau
Zoom Test for online presenters	14:00-15:00	<u>ZOOM ID: 880 7857 6196</u>

Zoom Test Timetable

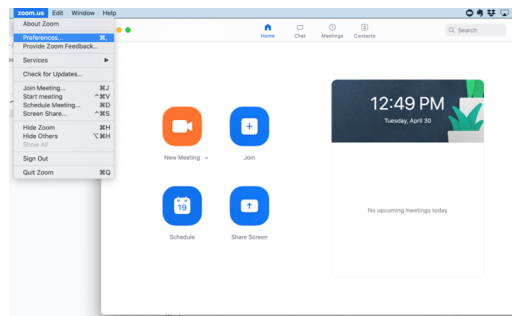
- ✧ Participants who are going to do an online presentation are required to join the rehearsal in Zoom on Friday, January 12, 2024. Duration: 3min apiece. Feel free to leave after you finish the test.
- ✧ We will test control panel including screen sharing, audio, video and "Raise Hand" feature, etc. Please get your presentation slides and computer equipment prepared beforehand.


14:00-14:30	V025 V017 V030 V028 V1001 V024 V031 V022 V010
14:30-15:00	Alternative time for participants who are unavailable at allocated time. Other online participants, includes but not limited to keynote speaker, session chair, committee member, listener.


Zoom Guidance

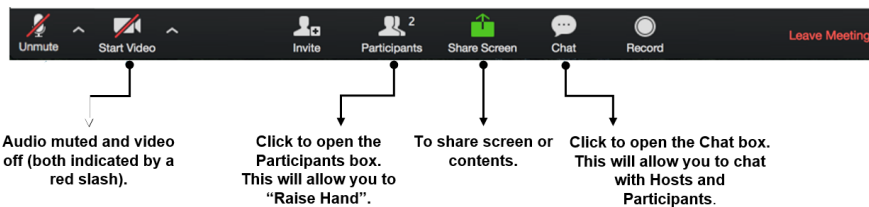
 You can join the meeting without sign-in process. Just put the meeting ID and join us.

 URL: <https://zoom.us/download>



 Each meeting has a unique 9, 10, or 11-digit number called a **meeting ID** that will be required to join a Zoom meeting.

 For any questions on the meeting day, you can text privately to "Assistant" for help.



Audio muted and video off (both indicated by a red slash).

Click to open the Participants box. This will allow you to "Raise Hand".

To share screen or contents.

Click to open the Chat box. This will allow you to chat with Hosts and Participants.

Saturday, January 13, 2024

Keynote Speech**Vitoria III, 2nd floor**
ZOOM ID: 880 7857 6196**Chairman:** Prof. Hui Zhang, BNU-HKBU United International College, China**9:00-9:10 Opening Remarks**Prof. Chi Man Pun, University of Macau, Macau, China**9:10-9:55 Keynote Speech I: Enhancing Language Models through Improved Pre-Training and Fine-Tuning**Prof. James Tin-Yau Kwok (IEEE Fellow), The Hong Kong University of Science and Technology, Hong Kong, China**9:55-10:40 Keynote Speech II: Clothed Human Model Reconstruction and Generation**Prof. Kenneth K.Y. Wong, University of Hong Kong, Hong Kong, China10:40-11:10 Group Photo & Coffee Break (2nd Floor)**11:10-11:55 Keynote Speech III: Multiple Ground Target Tracking with a Drone**Prof. Seokwon Yeom, Daegu University Gyeongsan, South Korea

12:00-13:30 Lunch (M Floor)

Parallel Session (Onsite)**Onsite Session 1A (13:30-15:15)****Vitoria III, 2nd floor****Data Communication and Image Processing Technology**

V001-A V007 V009 V002-A V011 V012 V014-A

15:15-15:45 Coffee Break (2nd Floor)**Onsite Session 1B (15:45-17:15)****Vitoria III, 2nd floor****Data Communication and Image Processing Technology**

V021 V403-A V027 V404-A V029 V406-A

18:00-20:00 Dinner
CATALPA GARDEN 梓園, 2nd Floor**Parallel Session (Online)****Online Session 1 (13:30-15:45)****Room A: ZOOM ID: 88078576196****Image Detection and Processing**

V025 V017 V030 V028 V1001 V024 V022 V010 V031

Keynote Speaker I (UTC+8)

Saturday, January 13, 2024

9:10-9:55

Vitoria III, 2nd floor

ZOOM ID: [880 7857 6196](#)



[Prof. James Tin-Yau Kwok](#)

IEEE Fellow, The Hong Kong University of Science and Technology, Hong Kong, China

Speech Title: Enhancing Language Models through Improved Pre-Training and Fine-Tuning

Abstract: Language models (LMs) are essential in natural language processing and vision-language modeling. However, several challenges arise in pre-training and fine-tuning of LMs. First, when learning through unsupervised pre-training, information that are semantically irrelevant may negatively affect downstream tasks, leading to negative transfer. Second, cross-modal masked language modeling is often used to learn vision-language associations in vision-language models. However, existing masking strategies may be insufficient in that the masked tokens can sometimes be simply recovered with only the language information, ignoring the visual inputs. Lastly, prompt tuning is effective in fine-tuning LMs on downstream tasks with limited labeled samples, but prompt design is difficult.

To tackle these issues, we propose several measures. First, we introduce a new pre-training method that trains each expert with only semantically relevant data through cluster-conditional gates. This allows downstream tasks be allocated to customized models pre-trained on data most similar to the downstream data. Second, on pre-training vision-language models, we use a masking strategy based on the saliencies of language tokens to the image. Lastly, we use meta-learning to learn an efficient prompt pool that can extract diverse knowledge from historical tasks. This allows instance-dependent prompts to be constructed from the pool without tuning the whole LM. Experimental results show that these measures can significantly improve the performance of LMs.

Dr. Kwok is a Professor in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. He received his B.Sc. degree in Electrical and Electronic Engineering from the University of Hong Kong and his PhD degree in computer science from the Hong Kong University of Science and Technology. Prof. Kwok served/is serving as an Associate Editor for the IEEE Transactions on Neural Networks and Learning Systems, Neural Networks, Neurocomputing, Artificial Intelligence Journal, International Journal of Data Science and Analytics, Editorial Board Member of Machine Learning, Governing Board Member and Vice President for Publications of the Asia Pacific Neural Network Society. He also served/is serving as Senior Area Chairs / Area Chairs of major machine learning / AI conferences including NeurIPS, ICML, ICLR, IJCAI, AAAI and ECML. He is recognized as the Most Influential Scholar Award Honorable Mention for "outstanding and vibrant contributions to the field of AAAI/IJCAI between 2009 and 2019". He is an IEEE Fellow.

Keynote Speaker II (UTC+8)

Saturday January 13, 2024

9:55-10:40

Vitoria III, 2nd floor

ZOOM ID: [880 7857 6196](https://us03zoom.us/j/88078576196)



[Prof. Kenneth K.Y. Wong](#)
University of Hong Kong, Hong Kong, China

Speech Title: Clothed Human Model Reconstruction and Generation

Abstract: The rapid development of the Metaverse, catalyzed by the advancements in virtual reality (VR) and augmented reality (AR) technologies, has resulted in a growing demand for 3D human models which are commonly used in applications like VR chat rooms and games. Historically, such 3D human models are created using CAD software, which is both time-consuming and expert-dependent. In this talk, we first introduce a flexible framework, named SeSDF, that can reconstruct a clothed human model from an arbitrary number of images under an uncalibrated setting. At the core of our framework is our novel self-evolved signed distance field module which allows the framework to learn to deform the signed distance field derived from a SMPL-X body model fitted to the image(s). We also introduce a simple method for self-calibration of multi-view images via the fitted SMPL-X parameters. This lifts the requirement of tedious manual calibration. In the second half of this talk, we introduce a generative framework, named DreamAvatar, that can generate high quality 3D human avatars with controllable poses based on text prompts. DreamAvatar utilizes a trainable NeRF for predicting density and color features for 3D points and a pre-trained text-to-image diffusion model for providing 2D self-supervision. Specifically, it leverages an SMPL-X model to provide rough pose and shape guidance for the generation. DreamAvatar is capable of generating avatars with detailed geometry and texture, establishing a new state-of-the-art for text-and-shape guided 3D human avatar generation.

Dr. Wong received the PhD degree in computer vision from the University of Cambridge in 2001. Since then, he has been with the Department of Computer Science at The University of Hong Kong, where he is now an Associate Professor. His research interests are in computer vision and machine intelligence. His research works include camera calibration, 3D reconstruction, image super-resolution, inpainting and restoration, and text-to-3D. He has published over 150 peer-reviewed journal and conference papers. Many of his works appear in top venues, including CVPR, ECCV, ICCV, TPAMI, IJCV, and TIP. He is currently an Associate Editor of IJCV.

Keynote Speaker III (UTC+8)

Saturday January 13, 2024

11:10-11:55

Vitoria III, 2nd floor

ZOOM ID: [880 7857 6196](#)



[Prof. Seokwon Yeom](#)
Daegu University Gyeongsan, South Korea

Speech Title: Multiple Ground Target Tracking with a Drone

Abstract: In this keynote speech, multiple ground target tracking with a small drone is introduced. The entire process consists of moving object detection and target tracking. Moving object detection consists of frame extraction and thresholding, morphological operations, and false alarm removing based on the size and shape of the object. Segmentation by k-means clustering is utilized in thermal video. Targeted tracking consists of the following steps: track initialization, measurement-trace association, state prediction and estimation, track-trace association, track termination, and validation testing. The measurement that is statistically nearest to the state prediction updates the target's state. With the improved track-to-track association, the fittest track is selected in the track validation region, and the direction of the displacement vector and velocity vectors of the two tracks are tested with an angular threshold. The coordinates of the image are converted to real-world based on the angular field of view, tilt angle, and altitude of the camera. In the experiments, various scenarios are tested including drone flight, bird-eye and oblique views, and thermal imaging. Tracking performance was evaluated by total track life (TTL) and mean track life (MTL). Promising results were obtained for 86 targets within approximately 1 km from the drone.

Seokwon Yeom has been a faculty member of Daegu University since 2007. He is now a full professor of the same university, School of AI. He has a Ph.D. in Electrical and Computer Engineering from the University of Connecticut in 2006.

His research interests are intelligent image/optical information processing, deep/machine learning, and target tracking. He has researched on multiple target tracking for the airborne early warning system, three-dimensional image processing with digital holography and integral imaging, photon-counting linear discriminant analysis and photon-counting nonlinear matched filter, millimeter wave and infrared image analysis, and long-distance target tracking for aerial surveillance and search and rescue mission with a small unmanned aerial vehicle.

He has been a guest editor of Applied Sciences and Drones in MDPI. He has served as a board member of the Korean Institute of Intelligent Systems, and a member of the board of directors of the Korean Institute of Convergence Signal Processing. He was program chair of ICCCS2015, ISIS2017, iFUZZY2018, ICCCS2019, ADIP2021-2023, IPMV2024. He was a vice director of the AI homecare center and a head of the department of IT convergence engineering at Daegu University in 2020, a visiting scholar at the University of Maryland in 2014, and a director of the Gyeongbuk techno-park specialization center in 2013.

Onsite Session 1A (UTC+8)

Saturday January 13, 2024

13:30-15:15

Vitoria III, 2nd floor

Basic Data Communication and Image Processing Technology

Chairperson: *Asst. Prof. Kai Han, The University of Hong Kong, Hong Kong, China*

<p>V001-A 13:30-13:45</p>	<p>Joint Detection: simultaneous achieving underwater image object detection and image enhancement Chao Yang, Southeast University, China</p> <p>Abstract-Underwater images are often less accurate when used for object detection due to the poor quality of underwater images. Most of the currently available solutions involve pre-processing, such as image enhancement, to detect objects, however this often leads to an increase in computational resource usage and the pre-processing is often ineffective. Therefore, this paper innovatively combines these two efforts and proposes joint detection network (JD detection). Through this generation process of underwater image enhancement, object detection shares clean features, thus facilitating underwater object detection, while the object detection results will guide the direction of underwater image enhancement. For the JD detection network, a shape prior enhancement module is proposed in this paper to enhance the network's modeling of global and local shape features. Through experimental validation, we show that our method combined with multiple object detection networks all achieve a remarkable performance gain and a significant improvement in processing speed.</p>
<p>V007 13:45-14:00</p>	<p>Advanced Multi-Dimensional Feature Fusion Attention Module Yi Shi, Northeastern University, China</p> <p>Abstract-Recent study has emphasized the importance of establish multidimensional information dependencies between weight vectors and input feature maps, in the process of calculating attention. However, although existing networks establish the connection from different perspectives, the connection presented is relatively limited, and the network's differentiation between important and non important information is insufficiency, which inevitably leads to effective information loss. This artical studies an efficient channel attention mechanism that can fuse multi-dimensional feature information, implement the interaction of channel and spatial position feature from both independent channels and global cross channels diemensions, and able to expand important information while suppress unimportant information. We propose the SW-SE block, which assigns the spatial position information of the cross channel to the process of calculating channel attention, strengthens information exchange between multiple channels, establishes closer connections, and obtains channel weight vectors with better expressiveness while greatly enhancing feature sampling ability. We have conducted ablation experiments on various mainstream network structures, and have achieved fine results in multiple aspects, e,g, classification, object detection and visualization. We reached 3.12% and 1.41% top-1 accuracy growth based on Resnet 50/100 on CIFAR10/100 respectively, and 4.01% on light weight network, along with 8.57% increased on AP_{75} for object dectection on PASCAL VOC2007/2012, with only a small number of parameters and computation time increased.</p>

<p>V009 14:00-14:15</p>	<p>Gesture Detection Using an Infrared Camera Majid Ahmadi, University of Windsor, Canada</p> <p>Abstract-Gestures and body language exert a significant impact on communication and interactions among individuals. This project's objective entailed the development of a gesture recognition system that analyzes the way people interact with their surroundings. The careful selection of the five recognized gestures ensures that the system remains focused on key limb placements that may dictate how one is perceived by the surrounding public. As the infrared (IR) camera captures real-time data, the system's ability to operate in low-light conditions further adds to its practicality and versatility in various environments. The convolutional neural network (CNN) plays a central role in the system's accuracy and efficiency. Its intensive training on a diverse dataset of images allows it to discern the distinct visual patterns associated with each gesture. As a result, the algorithm's Mean Average Precision (mAP) of 71.84% attests to its proficiency in accurately recognizing and classifying gestures.</p>
<p>V002-A 14:15-14:30</p>	<p>3D Reconstruction Algorithm of Human Whole-Body Skeleton Based on 2D-3D Image Registration Yuan Gao, Southeast University, China</p> <p>Abstract—In recent years, the use of imaging as an aid in orthopedic treatment has become a trend. The 3D CT images commonly used in orthopedic treatment take a long time to shoot and require a lot of radiation. It is very meaningful to study a method for rapid 3D reconstruction of human skeleton based on 2D X-ray images. In this study, the algorithm based on image registration is used to achieve fast 3D reconstruction based on real X-ray images and human bone standard membranes. The digital reconstruction projection (DRR) technology is used to project the standard 3D membrane body to obtain a 2D projection image. Image segmentation and image registration deep learning algorithms are applied to obtain the optimal rigid transformation matrix parameters for image registration between 2D projection images and real X-ray images. According to the registration transformation matrix, 3D transformation is performed on the standard phantom to obtain the 3D reconstruction result. In the experiments, human leg bones were used to evaluate the performance of the algorithm. The experimentally obtained registration accuracy is 2 mm and the average 3D reconstruction time is 11 seconds.</p>
<p>V011 14:30-14:45</p>	<p>Thai License Plate Recognition using SSD MobileNet and EasyOCR Janya Sainui, Prince of Songkla University, Thailand</p> <p>Abstract—Abstract-In this paper, we propose Thai license plate recognition for video utilizing deep learning approaches. Although many methods were proposed in the literature, most of them evaluated their performance using images not videos. Performing license plate recognition on videos is more challenging than that of images because moving cars cause blurred images, as well as incorrect and redundant number plates will be obtained. Therefore, our goal here is to automatically save the correct vehicle registration number with less redundancies in a database. To do so, we mainly use the SSD MobileNetV2 for license plate detection from video frames and then utilize EasyOCR for number plate recognition. Moreover, we introduce the post-processing process to filter out redundant and incorrect number plates before saving the number plates into the database. The experimental result of 100 images demonstrates that SSD MobileNetV2 obtains 99.00% accuracy for license plate detection, while EasyOCR achieves 92.00%-character accuracy for number plate recognition. The evaluation on videos shows that our proposed procedure achieves the license plate detection accuracy of 91.70%, the number plate recognition accuracy of 83.46%, and the character recognition accuracy of 96.99%.</p>

<p>V012 14:45-15:00</p>	<p>A Sorting Method of Irregular Feature Point Array in Camera Calibration Lemiao Yang, Beihang University, China</p> <p>Abstract—Camera calibration is an important factor affecting the precision of vision measurement, and the correct ordering of target featurpoints is the precondition of high-precision vision measurement. Most of the existing methods of feature point array sorting have good sorting effects for regular arrays, but with further research on high-precision vision measurement, the target of irregular feature point array with higher precision of feature points extraction or anti-noise ability is proposed gradually, and the sorting requirement of these irregular feature point arrays is difficult to meet. In order to achieve the correct sorting of irregular feature point arrays, this paper proposes a method for sorting the feature point arrays based on convex hull and projection matrix. The convex hull and convexity are used to identify and locate the corner points of the feature point array. The projection matrix maps the feature point array column to a non-rotational and non-translation array, which is sorted according to the established rules to realize the automatic ordering of the irregular feature point array. The effectiveness of the proposed method is verified by comparative experiments.</p>
<p>V014-A 15:00-15:15</p>	<p>IAFPN: Interlayer Enhancement and Multilayer Fusion Network for Object Detection Zhicheng Li, Southeast University, China</p> <p>Abstract—Abstract-Feature Pyramid Network (FPN) improves object detection performance by means of top-down multi-level feature fusion. However, the current FPN-based methods have not effectively utilized the interlayer features to suppress the aliasing effects in the process of feature downward fusion. We propose an interlayer attention feature pyramid network that attempts to integrate attention gates into FPN through interlayer enhancement to establish the correlation between context and model, thereby highlighting the salient region of each layer and suppressing the aliasing effects. Moreover, in order to avoid feature dilution in the process of feature downward fusion and inability of multilayer features to utilize each other, simplified non-local algorithm is used in the multilayer fusion module to fuse and enhance the multiscale features. A comprehensive analysis of MS COCO and PASCAL VOC benchmarks demonstrate that our network achieves precise object localization and also outperforms current FPN-based object detection algorithms.</p>

Onsite Session 1B (UTC+8)

Saturday January 13, 2024

15:45-17:15

Vitoria III, 2nd floor

Data Communication and Image Processing Technology

Chairperson: Prof. Majid Ahmadi, University of Windsor, Canada

<p>V021 15:45-16:00</p>	<p>Real-Time Predictability Analysis and Enhancement of Deep-Learning-Based Object Tracking Mingyang Zhang, KU Leuven, Belgium</p> <p>Abstract—While Siamese object tracking has witnessed significant advancements, its time-predictable implementation on embedded devices remains inadequately addressed. To bridge this gap, we propose to analyse the real-time predictability of components of a deep-learningbased object tracking system and then craft a dedicated hardware accelerator specifically for the bottleneck. This method seamlessly integrates advanced tracker features and propels the tracker’s speed and time-predictability on embedded systems. Implemented on a KV260 board, our quantized tracker demonstrates superior performance. These findings spotlight the immense promise of hardware acceleration in real-time object tracking and set a benchmark for forthcoming hardware-software co-design pursuits focused on achieving time-predictable object tracking.</p>
<p>V403-A 16:00-16:15</p>	<p>Clear cell renal cell carcinoma grading based on segment anything model Yunbo Gu, Southeast University, China</p> <p>Abstract—Nuclear grades are one of the most significant prognostic factors for clear cell renal cell carcinoma (ccRCC), early diagnosis and grading of ccRCC can assist clinicians in making individualized surgeries for patients, so it is necessary to develop a noninvasive model to predict the grades of ccRCC preoperatively. Inspired by the Segment Anything Model (SAM), a powerful foundation model that has achieved much attention for its impressive capability in 2D natural image segmentation, we propose SAM_ccRCC to utilize its generalization ability for ccRCC tumor grading. Firstly, considering the tumor heterogeneity, our model takes the whole 3D volumetric images as input, and processes it in a simple and effective way, rather than in a slice-by-slice mode from other SAM-based methods. Secondly, in order to integrate the medical domain-specific knowledge to our model and reduce the training parameters, we froze the original parameters of the SAM encoder and exploit an effective adaptation technique. Lastly, we combine the image features with other additional knowledge based on the contrastive learning representation to predict the tumor grades. We leverage a publicly available dataset to train and test our model, besides, a private dataset collected at our hospital is used as an external validation dataset for our model.</p>
<p>V027 16:15-16:30</p>	<p>Research on Image Segmentation Algorithm for Cellular Cortical Proteins Jianxin Zhang, Zhejiang University, China</p> <p>Abstract—Abstract-The proteins (such as RhoA and MYB) contained in the cell cortex play an important role in the cell mitosis process. In order to study the impact of cell cortex proteins on cell mitosis, it is necessary to extract and segment the contours of the cell cortex proteins. This study proposes a digital image processing method</p>

	<p>based on the OpenCV library to realize an automated method of extracting cell cortex contours and calculating brightness, and extracts the contours of cell cortex proteins from segmented single cell images. In the preprocessing part, adaptive histogram equalization and background removal based on distance transformation are used to increase the accuracy. It also uses the structural characteristics of the cell itself and combines the thickness of the cell cortex to divide the inner and outer contours, removing the synaptic interference caused by cell extrusion, making the accuracy reach 96%. In addition, this method also greatly improves the efficiency of cell cortex contour extraction, reducing the contour extraction time from at least 60 minutes to 2 minutes, which can greatly improve the efficiency of research.</p>
<p>V404-A 16:30-16:45</p>	<p>A Self-Supervised Spectral-Enhanced Network for Photon-Counting Spectral CT Imaging Qianyu Wu, Southeast University, China</p> <p>Abstract — Photon-counting detector (PCD) based computed tomography (CT) imaging faces great challenges. Various kinds of cross talk between detector elements and the limited number of photons in the narrow energy-bin may cause the reconstructed image to be degraded by severe noise. Although data-driven supervised learning methods have revealed great potential, PCD CT, as a new imaging modality, is difficult to collect sufficient clinical data for supervised training. To this end, this paper presents a self-supervised two-step training framework for PCD CT denoising without any reference. Specifically, we first implement self-supervised learning by constructing pseudo training pairs, and obtain coarse noise reduction results from the images reconstructed by traditional simultaneous iterative reconstruction technique. Then, we introduce a novel contrastive similarity regularization to incorporate noisy images and high-quality predictions into contrast learning, which ensures that the restored image is pushed far away from the noisy image in the representation space. In addition, we propose a spectral-enhanced network with a learnable low-rank dictionary to effectively exploit the spatial similarity of each specific energy bin and inter-channel correlation between energy bins. Extensive comparisons and analyses on both simulated and real data have demonstrated that the proposed method can achieve promising results for PCD CT image reconstruction in terms of noise suppression and contrast-to-noise ratio and is meaningful in theory and clinical practice.</p>
<p>V029 16:45-17:00</p>	<p>A Rolling Ball Segmentation and Pairing Algorithm Based on Cell Fluorescence Microscopic Stack Jiahao Chen, Zhejiang University, China</p> <p>Abstract—In order to explore the effect of a certain cell membrane protein (NMIIB and RhoA) on regulating the tension of the cortical layer, we segmented and paired the cells in the fluorescence microscopic stack of membrane proteins of two groups cells in different states (natural or compressed), to compare the intensity changes of the membrane proteins of the same cell before and after compressed. We propose a method for cell segmentation and pairing based on cell fluorescence microscopy stacked images, using a slice by slice scanning method to compare the information of each cell in different slice to determine its intermediate layer; In single-slice segmentation, we adopted the rolling ball algorithm to strip the inner contour of the cell membrane based on the characteristics of its membrane protein image, effectively avoiding the impact of cell adhesion; Finally, pairing factors were calculated based on the shape characteristics of each cell in the two images for pairing, taking into account the impact of overall field of view migration.</p>

V406-A
17:00-17:15

Resilient Base Station Parameter Optimization amidst UE Parameter Anomalies: A Strategic Zeroing-Out Approach

Sudharshan Paindi Jayakumar, Nokia Bell Labs, France

Abstract—The accurate prediction of Base Station (BS) parameters based on User Equipment (UE) features is critical for optimizing wireless communication networks. However, anomalies, inconsistencies, or missing data within UE parameters often pose significant challenges, impeding the precision of predictive models for BS parameter optimization. To address this pressing issue, this study introduces a comprehensive and intricate Zeroing-Out Strategy specifically tailored to handle UE parameters in the context of predicting optimized BS parameters. The proposed strategy consists of a meticulously designed two-phase methodology. Initially, sophisticated anomaly detection techniques are deployed to rigorously identify and subsequently remove unreliable or anomalous UE features from the dataset. Leveraging advanced anomaly detection algorithms, this phase ensures the removal of inconsistent or erroneous UE data points, providing a cleaner dataset for subsequent model training. Following the anomaly removal phase, a targeted and strategic training strategy is employed. During this phase, the model is trained with a deliberate focus on adaptability to missing or removed UE features. Individual UE parameters are selectively withheld during model training while predicting BS parameters, simulating scenarios where specific UE data might be absent or unreliable. This strategic approach compels the model to learn and reconstruct missing or removed UE data, thereby enhancing its resilience and adaptability to unpredictability in the UE data space. The primary aim of this strategic zeroing-out approach is to enable accurate predictions of optimized BS parameters even when certain UE features are missing or removed due to anomalies. Through extensive experimental validation using real-world datasets, the effectiveness and robustness of the proposed strategy are rigorously evaluated. Results showcase a marked improvement in prediction accuracy, demonstrating the strategy's efficacy in mitigating the impact of missing or anomalous UE features on BS parameter predictions. By bridging the gap between anomaly detection and model adaptability, this innovative strategy presents a promising solution to enhance the reliability and efficiency of wireless communication systems' parameter optimization. The insights and findings from this research not only contribute to advancing predictive modeling in wireless networks but also hold significant promise for improving the resilience of network optimization processes in real-world deployment scenarios.

Online Session 1 (UTC+8)

Saturday, January 13, 2024

13:30-15:45

Room: [ZOOM ID: 880 7857 6196](#)

Image Detection and Processing

Chairperson: Prof. Dr. Yew Kee WONG Eric, Hong Kong Chu Hai College, Hong Kong, China

<p>V025 13:30-13:45</p>	<p>Dual-Channel Reliable Breast Ultrasound Image Classification Based on Explainable Attribution and Uncertainty Quantification Shuge Lei, University of South Carolina, USA</p> <p>Abstract—This paper focuses on the classification task of breast ultrasound images and researches on the reliability measurement of classification results. We proposed a dual-channel evaluation framework based on the proposed inference reliability and predictive reliability scores. For the inference reliability evaluation, human-aligned and doctor-agreed inference rationales based on the improved feature attribution algorithm SP-RISA are gracefully applied. Uncertainty quantification is used to evaluate the predictive reliability via the Test Time Enhancement. The effectiveness of this reliability evaluation framework has been verified on our breast ultrasound clinical dataset YBUS, and its robustness is verified on the public dataset BUSI. The expected calibration errors on both datasets are significantly lower than traditional evaluation methods, which proves the effectiveness of our proposed reliability measurement.</p>
<p>V017 13:45-14:00</p>	<p>Unsupervised Multi-Source Domain Adaptation for Pedestrian Re-identification : A Study in Noise-Resilient Learning Xia Liu, Beijing Institute of Technology, Zhuhai, China</p> <p>Abstract—Pedestrian re-identification is a technology used for cross-camera pedestrian retrieval in surveillance data. In the context of massive data, unsupervised domain adaptation methods have gained increasing attention. Currently, one prevalent approach during training on unlabeled target domains involves generating pseudo-labels. However, due to the inability to guarantee the correctness of pseudo-labels, they inevitably carry noise, subsequently affecting the network's training process. In this paper, we propose a novel approach that better constrains noise and demonstrates strong performance. First, building upon the DBSCAN clustering method, we generate soft labels for samples, thereby providing fine-grained category loss supervision. Specifically, we introduce an adaptive reverse cross-entropy loss to impose adaptive constraints on the pseudo-label generation process. We conduct various experiments on four datasets across three transfer tasks, and the results consistently highlight the superiority of our method, providing strong evidence for its effectiveness.</p>
<p>V030 14:00-14:15</p>	<p>Generative Adversarial Model Equipped with Contrastive Learning in Map Synthesis Arpan Mahara, Florida International University, United States</p> <p>Abstract—In the dynamic field of urban planning and the context of unprecedented natural events, such as hurricanes, the fast generation of accurate maps from satellite imagery is paramount. While several studies have utilized Generative Adversarial Networks (GANs) for map generation from satellite images, the present work introduces a new approach by integrating contrastive learning into the GAN framework for enhanced map synthesis. Our methodology distinctively employs positive sampling by aligning similar features (e.g., roads) in both satellite images and their corresponding map outputs, and contrasts this with negative samples for</p>

	<p>disparate elements. This approach effectively replaces the conventional cyclic process in GANs with a more streamlined, unidirectional procedure, leading to improvements in both the quality of the synthesized maps and computational efficiency. We show the effectiveness of our proposed model, offering an advancement in map generation for remote sensing applications.</p>
<p>V028 14:15-14:30</p>	<p>Vehicle Detection Based on YOLOv7 for Drone Aerial Visible and Infrared Images Zhou Tao, Northwestern Polytechnical University, China</p> <p>Abstract—Object detection using drone-captured aerial images holds great significance for real-time traffic detection and control by traffic management authorities. This paper introduces a novel algorithm based on the widely-used YOLOv7 model. The algorithm presents an enhanced evaluation metric, the Normalized Gaussian Wasserstein Distance (NWD), utilizing normalized Wasserstein distance for small object detection. NWD measures object similarity by considering associated Gaussian distributions. To address challenges from complex backgrounds and redundant features due to irrelevant noise in aerial images, the paper introduces the CBAM attention mechanism. This mechanism improves feature expression for vehicle detection, enabling selective focus on target regions. The proposed algorithm is evaluated on the Drone Vehicle dataset and compared with state-of-the-art algorithms. Experimental results demonstrate favorable average precision values in both visible light and infrared images. The improved network model leads to a 1.2% increase in mean average precision (mAP) for visible light image detection and a 0.9% increase for infrared image detection.</p>
<p>V1001 14:30-14:45</p>	<p>Inpainting of coherent regions in binary images based on Coherence Enhancing Cahn-Hilliard equation Rajrup Banerjee, Jadavpur University, India</p> <p>Abstract—The act of interpolating missing or damaged portions of an image by applying information from the neighborhood is known as image inpainting. A lot of fourth order Partial Differential Equation (PDE) based models have been devised to solve the inpainting problem, but most do not generate satisfactory results when solving the problem in a region containing coherent lines. In this paper, we present a new fourth order PDE for solving the image inpainting problem in a region containing coherent structures in binary or black and white images. Our method is based on Cahn-Hilliard Equation coupled with the coherence enhancing structure tensor to address the inpainting in the regions containing coherent structures. The coherence enhancing structure tensor has been extensively used by J. Weickert in his formulation of anisotropic diffusion equation. It works on the principle of enhanced diffusion in homogenous regions and prohibited diffusion across edges of homogenous structures even if the edges are not along the X or Y axes. We have proposed an explicit-implicit numerical scheme based on splitting two convex energy terms. Then the proposed PDE is solved using Fourier spectral method. From the results, we can see that proposed method generates far better inpainted images than the state-of-the-art methods.</p>
<p>V024 14:45-15:00</p>	<p>Imaging Technology in High-speed Dynamic Scene and Its Application in the Dynamic Detection on Pantograph-catenary Interaction Yan-guo Wang, Institute of Infrastructure Inspection, China Academy of Railway Sciences Corporation Limited, China</p> <p>Abstract—The interaction of pantograph-catenary is an important interaction relationship between vehicles and railway infrastructure, which determines the</p>

	<p>stability of power supply in high-speed railways. Dynamic performance of pantograph-catenary has to be checked regularly, in which the technology of machine vision plays an important role. Compared with video surveillance and other traditional industrial applications, there are some special technical problems in the dynamic detection of pantograph-catenary interaction, such as huge difference on outdoor light environments, rapid change of image background, and serious motion blur at high speed. In this paper, we proposed a method of visual inspection on pantograph-catenary interaction, based on technologies of high-frequency synchronous lighting and wide dynamic range imaging, and a vision system was developed. The advantages of the proposed method have been showed in experimental test, and the image quality has been significantly improved in high-speed dynamic scene, which shows great potential in visual data analysis in the dynamic detection on pantograph-catenary interaction.</p>
<p>V022 15:00-15:15</p>	<p>The Use of Virtual Reality Platforms to Improve Students' Speaking Skills A. Rahman, Vistula University: Akademia Finansow i Biznesu Vistula, Warsaw, Poland</p> <p>Abstract—This study explores the efficacy of virtual reality (VR) platforms in enhancing students' speaking abilities and addressing challenges in foreign language learning. Employing a mixed-methods approach, the research utilizes in-depth interviews and descriptive statistics to comprehensively assess the impact of a Virtual Reality-based SpeakEasy VR application. Qualitative insights from interviews inform quantitative inquiries, revealing students' positive perceptions and significant improvements in English speaking skills. Findings indicate high agreement across ten indicators (55.55% to 83.33%), confirming widespread support for the VR platform. Paired t-test results (t-value: 34.855) surpassing the critical threshold (2.7423) further affirm the platform's substantial positive impact. This research provides a nuanced understanding of VR platforms' influence on language skills, emphasizing their potential in educational contexts.</p>
<p>V010 15:15-15:30</p>	<p>Hybrid Color Watermarking Technique with Arnold Scrambling Sanjeeb Prasad Panday, Institute of Engineering, Tribhuvan University, Nepal</p> <p>Digital watermarking is one of the best techniques to protect the digital data from unauthorized access and provide copyright protection. Imperceptibility and robustness are the important features of digital watermarking. Color image watermarking method based on discrete wavelet transform (DWT), all phase discrete cosine bi-orthogonal transform (APDCBT) and singular value decomposition (SVD) is proposed in this paper. In this technique we hide a color image watermark into a color cover image balancing both imperceptibility and robustness. The three level DWT is applied to the Luminance (Y) channel of the color cover image. Block based APDCBT is applied to the resultant diagonal details (HH) sub band and the DC coefficient matrix is generated for SVD which then embeds the Red channel of the watermark image after implementing security with Arnold Transform followed by DWT and SVD. APDCBT transform is used instead of conventional DCT to reduce the block artifacts at low bit rates. The main reason to use the DC coefficient after APDCBT is to enhance the watermark imperceptibility. Furthermore, the proposed method achieved high robustness in resisting various signal processing and geometric attacks. This is justified by comparison with the reference watermarking method that doesn't implement the Arnold transform. Comparatively, the proposed algorithm has the better robustness for the most of the single attack as well as the mixed attack cases than the reference method.</p>

V031 15:30-15:45	<p>A Comparative Study of Swin-Based Enhanced Remote Sensing Image Classifications Berfin Kurtoglu, Harran University, Turkey</p> <p>In image classification methods, the quality of the input image plays an important role in improving classification performance. However, sometimes the low resolution and sharpness of remote sensing images can cause various problems in image analysis. Therefore, improving and correcting the quality of remote sensing images is of great significance for remote sensing image classification. In this study, five man-made and five natural field images were selected from the RSI-Cb remote sensing dataset, and the corresponding images were super-resolved by using the Swin-based HST, Swin2SR, and SwinIR transformers. The classifications were performed using the pre-trained architectures DenseNet121, Xception and EfficientV2_B3 and their performance was compared. The result of the experiments was that the classification of images was improved by applying super-resolution methods. It was found that better results were achieved especially for images of natural areas. The best result in terms of classification performance was obtained when the classification was performed with the Xception architecture by increasing the super-resolution results of the images with the HST algorithm, so that the classification accuracy increased from 99.18% to 99.59%.</p>
---------------------	--

Note
